



# Increasing Metagenomic Resolution of Microbiome Interactions Through Functional Phylogenomics and Bacterial Sub-Communities

Angélica Cibrián-Jaramillo\* and Francisco Barona-Gómez\*

Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), Unidad de Genómica Avanzada, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Irapuato, Mexico

## OPEN ACCESS

### Edited by:

Alejandro Sanchez-Flores,  
Universidad Nacional Autónoma de  
México, Mexico

### Reviewed by:

Veljo Kisand,  
University of Tartu, Estonia  
Sergios-Orestis Kolokotronis,  
Fordham University, USA

### \*Correspondence:

Angélica Cibrián-Jaramillo  
[acibrian@langebio.cinvestav.mx](mailto:acibrian@langebio.cinvestav.mx);  
Francisco Barona-Gómez  
[fbarona@langebio.cinvestav.mx](mailto:fbarona@langebio.cinvestav.mx)

### Specialty section:

This article was submitted to  
Bioinformatics and Computational  
Biology,  
a section of the journal  
Frontiers in Genetics

**Received:** 30 June 2015

**Accepted:** 17 January 2016

**Published:** 10 February 2016

### Citation:

Cibrián-Jaramillo A  
and Barona-Gómez F (2016)  
Increasing Metagenomic Resolution  
of Microbiome Interactions Through  
Functional Phylogenomics  
and Bacterial Sub-Communities.  
*Front. Genet.* 7:4.  
doi: 10.3389/fgene.2016.00004

The genomic composition of the microbiome and its relationship with the environment is an exciting open question in biology. Metagenomics is a useful tool in the discovery of previously unknown taxa, but its use to understand the functional and ecological capacities of the microbiome is limited until taxonomy and function are understood in the context of the community. We suggest that this can be achieved using a combined functional phylogenomics and co-culture-based experimental strategy that can increase our capacity to measure sub-community interactions. Functional phylogenomics can identify and partition the genome such that hidden gene functions and gene clusters with unique evolutionary signals are revealed. We can test these phylogenomic predictions using an experimental model based on sub-community populations that represent a subset of the diversity directly obtained from environmental samples. These populations increase the detection of mechanisms that drive functional forces in the assembly of the microbiome, in particular the role of metabolites from key taxa in community interactions. Our combined approach leverages the potential of metagenomics to address biological questions from ecological systems.

**Keywords:** microbiome, metagenomics, functional phylogenomics, co-culture, bacterial interactions, specialized metabolites

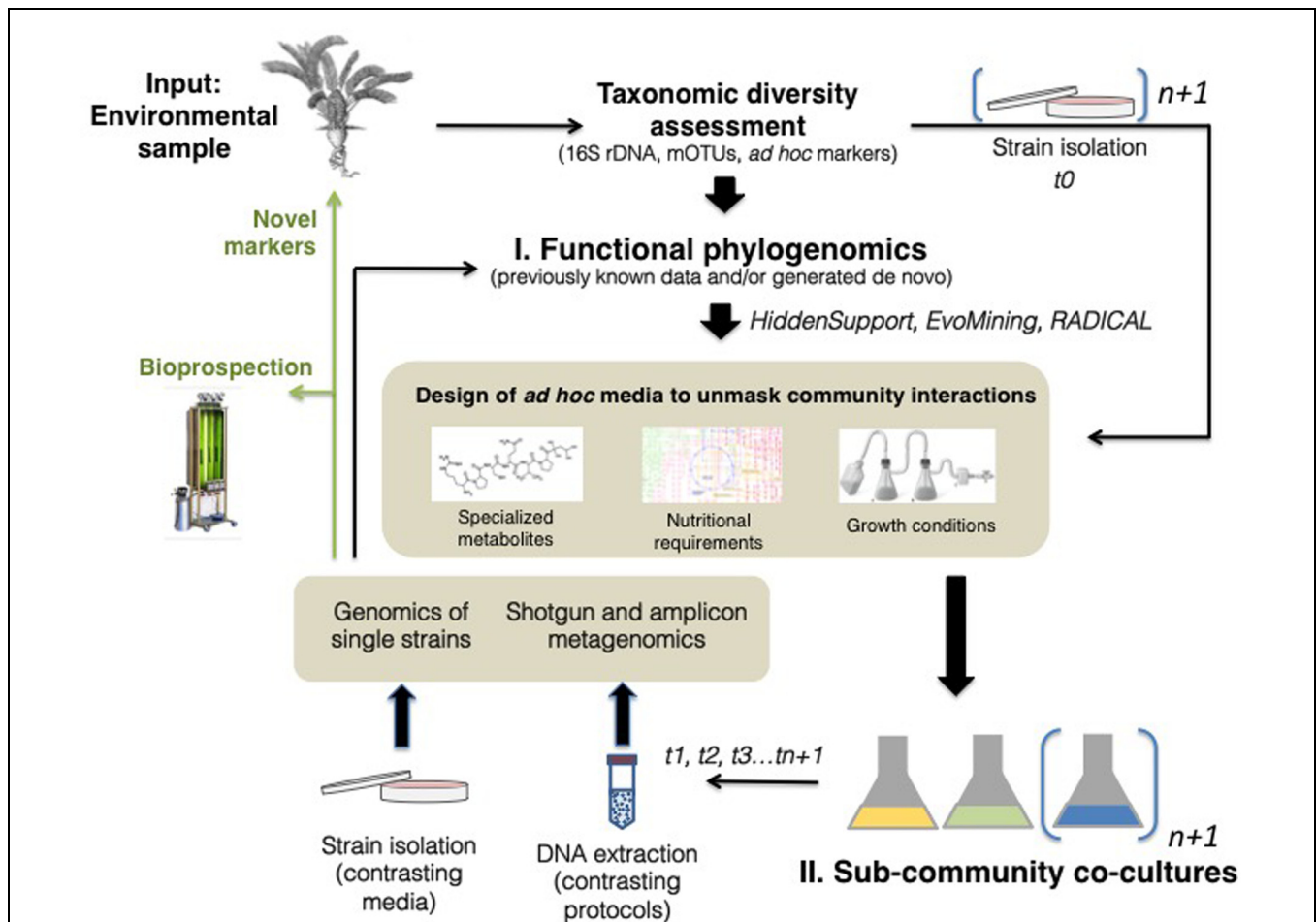
## INTRODUCTION

Biological understanding of the microbiome, defined as the set of microorganisms and their genomes in a particular environment (Boon et al., 2014), is one of the most exciting frontiers in science. High-throughput sequencing of single markers (16S rRNA gene) and shotgun metagenomics are now commonly used to describe the microbiome, revealing the presence of novel taxa (Wu et al., 2011; Lok, 2015); and increasing our understanding of the intimate interactions between symbiotic microbiomes and their hosts, in what has been termed the holobiont (Wilson et al., 2014). In the holobiont, the microbiome functions as an extended pheno/genotype of the host (Bordenstein and Theis, 2015), their rapid generation times enhancing the host's ability to adapt quickly, and likely providing adaptive metabolites or the enzymatic machinery to produce them (Rosenberg et al., 2010). Indeed, the microbiome and its host likely act as a single evolutionary unit, of which we know very little of.

A major challenge to understand the host–microbiome interaction and the holobiont with its environment, is to define the intersection between taxonomic and functional diversity of the microbiome. That is, who are the members of the microbial community, who is merely co-existing, who is interacting and how; and how this could impact the evolution of the holobiont. To date, most metagenomic approaches on their own, lack sequencing depth to identify taxonomic groups, assemble and annotate functional pathways, but most importantly, recover ecologically key taxa and their associated metabolites that could identify bacterial interactions within the community.

We are mainly interested in increasing our ability to provide evolutionary and functionally relevant information

of the microbiome community. We suggest a combined use of phylogenomic predictions and sub-community co-cultures in which biodiversity and their interactions are better understood. In particular, we are concerned with how to simplify the search for genetic information of functional importance from metagenomes, and how to simplify the composition of the microbiome enough that functional annotation can be interpreted in the context of bacteria–bacteria and bacteria–host interactions. One product of this combined approach is to generate functional information (e.g., gene clusters) that is presumed to have ecological and evolutionary relevance, and can therefore be used to validate biological hypotheses, by characterizing them back in an environmental sample or by downstream experimentation (Figure 1).



**FIGURE 1 | A combined phylogenomics and experimental approach to enhance measurement of bacterial interactions within the microbiome.**

Starting with an environmental sample, it would be possible to describe its taxonomic diversity, and use this information as well as published information (e.g. related taxa), to carry out phylogenomic analyses such as Hidden Support, EvoMining or RADICAL (see text for details). The main outcome (bold arrow) of the Functional Phylogenomics step is to identify genes with unique evolutionary and functional signatures that can be used to design *ad hoc* media co-cultures, in the form of information related to specialized metabolites, nutritional requirements or growth conditions. Different sub-community co-cultures can then be designed, and these could be measured at different time points, from days to years ( $t_1, t_2, t_3, \dots$ ). Contrasting DNA extraction techniques from co-cultures would help reveal various types of taxonomic groups, measured with metagenomic strategies. Likewise, strain isolation could be carried out to sequence individual genomes in-depth. The main outcome of the sub-community cultures is information (genomes, genes) that can be annotated in-depth in the context of a simplified but natural community. This information can be used to re-design the *ad hoc* media into new sub-communities that target specific groups, used as part of bioprospecting strategies, or as novel markers that can be used to validate the presence of the community members in the original environmental sample.

We propose two complementary strategies to achieve the goal of identifying and using functional and evolutionary meaningful information to understand the microbiome community. First, the use of phylogenomic tools to identify gene clusters or gene families with a distinct evolutionary pattern that can be used to infer taxonomic patterns and functional roles in the bacterial community. The main advantage of this computational approach is that these candidate genes constitute a reduced universe of evolutionary hypothesis based on function, which can be tested experimentally and/or with down-stream bioinformatics, rather than experimenting more broadly in the entire genome or metagenomic dataset. It differs from current efforts to determine phylogenetic diversity focused on 16S rRNA gene or markers with phylogenetic signal, such as metagenomic operational taxonomic units (mOTUs) (Sunagawa et al., 2013); or profiling methods based on many markers (Nguyen et al., 2014). It complements methods based in characterizing metacommunity of genes with some functional relevance (Boon et al., 2014), and methods that attempt to identify functional genes in the context of partitioning bacterial genomes into pan/shell genomes (Shi and Falkowski, 2008; van Tonder et al., 2014) or using information theory to quantify the degree of conflict or incongruence calculated from different types of data (Salichos et al., 2014), although we argue that a finer partition beyond a core and pangenome is required to be able to identify functional genes, perhaps more similar to GO-based phylogenetic classifications (Chai et al., 2014). It would also be possible to use metatranscriptomic profiling to help narrow down gene candidates of adaptive value based on their differential expression (Franzosa et al., 2014).

Second, we suggest that it is possible to sample sub-communities of the original environmental sample using co-culture strategies, under the assumption that functional interactions mediated by genes and their products will be easier to detect in more simple, pre-conceived functionally driven culture conditions. A distinctive advantage of the co-cultures over traditional microcosms whose reproducibility has been recently questioned (see e.g., Langenheder et al., 2006; Pagaling et al., 2014), is that interactions within a complex system may be better resolved by dissecting it onto sub-communities based on functional concepts. Functionally important genes identified with phylogenomic strategies can be used to guide the design of the co-cultures themselves, in the form of metabolites that can drive community interactions, nutrients required or even biological growth conditions (**Figure 1**). Co-cultures in themselves enable testing of directed hypotheses, and can generate data that can be used as a source of novel *ad hoc* markers, which can be validated in the original biological sample and used to re-design new co-cultures that enhance our ability to understand bacteria interactions (**Figure 1**).

The implications of knowing key functional players and their interactions in the microbiome enlightens evolutionary biology, but can also help solve two major issues in bioprospecting of secondary or specialized metabolites (Charlop-Powers et al., 2015; Ling et al., 2015), commonly known as secondary metabolites or natural products. That is, the discovery of genes encoding for novel enzymes; and ‘turning on’ biosynthetic genes directing the synthesis of specialized metabolites in ways that can

be adopted by synthetic biology approaches (Luo et al., 2015; **Figure 1**), facilitating the transition from pattern descriptions into deciphering mechanisms at different levels (Waldor et al., 2015).

## PHYLOGENOMICS TO DETECT AND CLASSIFY GENES OF FUNCTIONAL IMPORTANCE IN THE MICROBIOME

We describe three strategies aimed at identifying genes that have evolutionary and ecological relevance, and identifying hidden functional diversity within metagenomes in the context of a phylogeny, based on (a) revealing hidden support of genes within a concatenated alignment, as the alignment matrix resulting from the concatenation of all gene/protein partitions that are orthologous among the targeted genomes; (b) measuring the emergence of specialized metabolism in a topology to discover hidden chemical diversity; and (c) identifying clades and gene categories with incongruent evolutionary signals that suggest horizontal gene transfer (HGT), genome streamlining, or unique evolutionary trajectories (e.g., Shi and Falkowski, 2008). In combination, these phylogenomic methods can reveal ‘functional units’ that can be identified and classified, also helping to resolve species delimitations and their phylogeny, although this is not our main goal.

### (a) Hidden Support

Work by one of the co-authors (Cibrián-Jaramillo et al., 2010; Lee et al., 2011) previously developed a phylogenomic approach to identify genes of functional importance in plants. We suggest that this approach can be transferred to bacterial metagenomes and genomes to identify genes that have different functional roles, and biological processes putatively involved in species diversification. In this method, the authors provide a way to study the behavior of genes used to reconstruct a phylogeny through analysis of their effect on tree topology and branch support (Cibrián-Jaramillo et al., 2010). This approach is fundamentally different than classical phylogenetic analysis because the search for both orthologs and candidate genes is conducted with a phylogeny. If one assumes that the tree obtained from concatenated analysis using these orthologs best represents the evolutionary history of the taxa involved, then phylogenetic incongruence between a partitioned functional class of genes and the organismal phylogeny would suggest that the partition (gene) has experienced a unique evolutionary history relative to the organisms. These genes are essentially a set of hypotheses of functions and evolutionary mechanisms that can be validated experimentally.

### (b) EvoMining

Cruz-Morales et al.<sup>1</sup> developed a functional phylogenomics platform to identify expanded, repurposed enzyme families, which have been recruited from central metabolism into the context of specialized metabolism. EvoMining uses single-gene

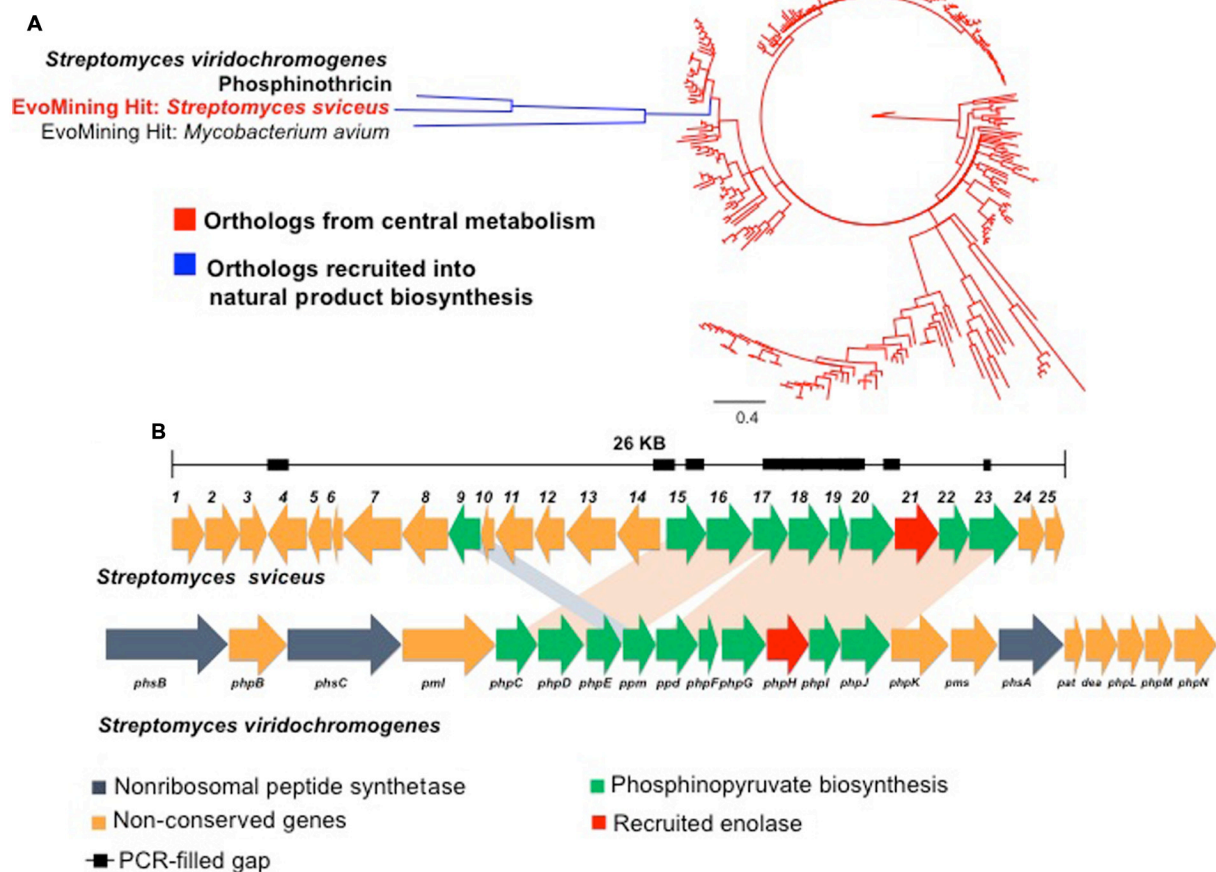
<sup>1</sup><http://www.biorxiv.org/content/early/2015/06/08/020503.abstract>

topologies to reveal characteristic phylogenetic signatures of rapid evolution that can be further analyzed in detail with the construction of hidden Markov models (HMM) profiles. These mechanisms leave a phylogenetic signature that can be measured in a topology by means of both clade formation and sequence divergence. Functional annotation of the phylogenetic tree, on the basis of genome-mining approaches (e.g., antiSMASH; Medema et al., 2011), provides validation and insights for visual inspection of the data. As this approach does not rely on sequence similarity searches of previously identified enzymes, but rather on recapitulation of an enzyme evolutionary process, it is less sensitive to missing data, which is relevant for low quality draft genomes (e.g., metagenomes). Subsequently, functional specificities of identified enzymes can be explored (Noda-García et al., 2015). An example of how EvoMining can lead to hidden signals in the context of metagenomics is provided in Figure 2. We show an EvoMining hit in the enolase enzyme family found in the highly fragmented genome of *Streptomyces sviveus* (1 scaffold

of 9 Mbp with 552 gaps and 8X coverage, GI: 297196766). Its contig containing a recruited enolase (GI: 297146550) typically involved in glycolysis, had 6 gaps including missing sequence at its 5' end. After closing gaps (PCR), the complete sequences for several phosphonate-related enzymes, namely, alcohol dehydrogenase (*phpC*), phosphonopyruvate decarboxylase (*ppd*), nicotinamide mononucleotide adenylyl transferase (*phpF*), carboxy-phosphoenolpyruvate synthase (*phpH*, EvoMining hit), and aldehyde dehydrogenase (*phpI*), could be annotated. Further sequence analysis suggested that indeed this locus encodes for a putative phosphinate biosynthetic gene cluster related to the phosphonic tripeptide phosphinothricin. This was recently confirmed by in-depth comparative genomics analyses (Blodgett et al., 2015; Figure 2).

### (c) RADICAL

The main goal of RADICAL is to reveal aspects of phylogenetic behavior otherwise not evident in individual gene trees or 'total



**FIGURE 2 | EvoMining example. (A)** Phylogenetic reconstruction of enolases extracted from an *Actinobacteria* genome database. After manual annotation, homologs believed to be involved in glycolysis were labeled in red. The clade predicted to include enolase homologs involved in natural product biosynthesis, including that from *S. sviveus* (EvoMining hit), is shown in blue. **(B)** Structural organization of the natural product biosynthetic gene cluster identified using EvoMining on the genome of *S. sviveus* (top). After manual closure of this locus after sequencing of PCR products (black boxes) the emerging cluster was annotated and compared with the organization of the biosynthetic gene cluster known to direct the synthesis of phosphinothricin in *S. viridochromogenes* (bottom) (Blodgett et al., 2015).



evidence' trees (Narechania et al., 2012). RADICAL sequentially concatenates randomly chosen gene partitions starting with a single-gene partition and ending with the entire genomic data set, with a tree built for every successive addition; until a library of trees is generated representing a large variety of differently sized random gene partitions. RADICAL analyses in cyanobacteria using gene subgroups confirmed that there are elevated levels of incongruence for genes involved in metabolism, but also showed that the distinction of 'core' and 'pan/shell' genome is not clear-cut. Furthermore, metabolism genes were the only functional class of genes supporting the monophyly of *Prochlorococcus*, suggesting that if indeed metabolism genes are affected by HGT more than other types of genes, there is virtually no support for a cohesive grouping of species within this clade (Narechania et al., 2012). Support for monophyletic groups can be used to identify which genes are congruent with species delimitations based on 16S rRNA genes or highly conserved markers, including MOTUs (Sunagawa et al., 2013) to distinguish phylogenetic history from functions due to community interactions.

Hidden Support and RADICAL are based on concatenation and topological congruence (Rokas et al., 2003; Gatesy, 2005). In essence, metagenomic data would be partitioned into evolutionary units independent of a core and pan-genome category, to reveal any genes that have different phylogenetic histories, but also those that affect the topology as a result of biased, missing data, or errors in annotation. The partition of metagenomes into evolutionary units is not a trivial task but it would be possible to deconvolute genes and identify orthologs from metagenomic data using binning methods that recover discrete units from metagenomic sequences (Mackelprang et al., 2011; Sharon et al., 2013; Alneberg et al., 2014); or by calculating gene abundances across various samples and correlating it with taxonomic distribution, such that we have taxon-specific profiles (Carr et al., 2013). A phylogeny-driven Bayesian test for the presence of an organism in a sample could also be done (Darling et al., 2014); or a combined strategy of ortholog clustering (e.g., OrthoMCL) followed by a phylogeny to identify and single out orthologs from paralogs (Chiu et al., 2006). Genes resulting from any of these methods of metagenomic extraction can be used to construct the trees required for Hidden Support and RADICAL.

In sum for this section, phylogenomic data would enable a *top-bottom* strategy in which we could predict metabolites from metagenomic sequences that are community interactors or mediators. We would be able to validate the role of keystone taxa through annotated gene clusters in bacterial metacommunities, in particular those that are essential for adaptive bacterial phenotypes or function as community interactors such as persister variants in antimicrobial resistance (Amato et al., 2014) or siderophore biosynthesis (D'Onofrio et al., 2010). It would be possible to develop markers (e.g., oligos based on the sequence of the identified functionally important genes) that can be amplified in an environmental sample or in an experimental culture as described below, to test their presence, distribution, and biological role in the microbiome (Figure 1).

## IDENTIFYING AND TESTING THE ROLE OF FUNCTIONAL INTERACTORS FROM METAGENOMES

The main question in this section is how do we define the interactions within a community in the context of the microbiome? We adopt the definition of community as 'multispecies assemblages, in which organisms live together in a contiguous environment and interact with each other' (Konopka, 2009). Early views of the community in ecology imply a tight interaction among its members (Clements, 1916) perhaps akin to the modern idea of the metacommunity in which there is dynamic movement of genes through microbial lineages (Boon et al., 2014), but more importantly, metabolic dependencies between community members (Hug et al., 2012). Thus, a community can be hypothesized as metabolic interactions that can be tested in the context of the microbiome's environment and/or its relationship with a host (i.e., holobiont).

The main technical issue, however, is to distinguish coexistence from a spectrum of interactions among the members of a set within a niche, viewed as 'a subset of those environmental conditions which affect a particular organism, where the average absolute fitness of individuals in a population is greater than or equal to one' (Kearney, 2006). Our main premise is that microbiome interactions can be described by experimentally testing and manipulating their role in the community assembly processes as it is done in eukaryote ecology (Fayle et al., 2015). In the bacterial microbiome, metabolites and the genes that underlie their presence and variation could be used as a proxy for interactions, including the identification of ecologically key organisms (Mills et al., 1993).

To achieve this we propose the use of functionally driven sub-community co-cultures, which are preceded by synthetic co-cultures typically involving arrays of isolates in the format of two strains at the time. In the best-case scenario, the latter may come from the same sample but not necessarily known to interact. Yet, co-cultures are increasingly used to understand general mechanisms by which bacteria may interact among themselves and with eukaryotic cells by means of metabolic exchange (Traxler and Kolter, 2015). The chemotypes sustaining such interactions can be assessed by high-resolution analytical approaches, such as imagining mass spectrometry directly from Petri dishes (Hoefer et al., 2012) and nuclear magnetic resonance (Wu et al., 2015). Even shallow sequencing coverage of these sub-populations would be enough to reach resolution of species diversity and the possibility to measure novel specialized metabolites directly in the context of the sub-community. In contrast with metagenomes obtained directly from environmental sampling, we do not aim to have thorough ecological representation. *Ad hoc* co-culture conditions are simply a proxy for the mechanistic niche in which we increase metagenomic resolution of a simplified metabolic niche.

Co-cultures of sub-communities starting from inoculum directly obtained from environmental samples, can be designed in various ways. For example, using substrates that enhance

diversity which are initially enriched for a particular bacterial genus or species suspected or shown to be prevalent in environmental samples. Additionally, sub-communities can be recovered from media that are as similar as possible to the original hosts or sites (Ling et al., 2015) or based on genome-scale metabolic network reconstructions from phylogenomic data, leading to nutritional requirements of key taxa (Barona-Gómez et al., 2012; **Figure 1**).

Thus, co-cultures which may be followed through various time series (days or years) would enable growth of typically rare species (Pagaling et al., 2014); genomic lineages with functional traits that are not easily recovered; or species that sustain metabolic exchanges with other bacterial symbionts. This could unmask genomic-lineages undergoing evolutionary processes, for example those driven by differential gene gain-and-loss. By genome sequencing of those strains, which may be recovered from the co-cultures, the resolution of the microbiome using metagenomics would be increased. In contrast with synthetic populations (Hoeftler et al., 2012), these sub-communities represent an experimental platform to reduce ecological complexity including at least some of the original members of the biological sample. They represent a complement to microcosms in that reproducibility and predictability, and thus biological resolution, is expected to increase in the co-cultures. Eventually we would have enough resolution to detect inactivated genes and pathway degeneration, as well as appearance of novel pathways, in certain niches that result from local adaptation (Hittinger et al., 2004), and as sequencing depth increases, metabolic networks in the context of the community could be reconstructed (Klitgord and Segrè, 2011).

Metagenomic data from sub-community populations also provides the possibility of interpreting species distributions with community ecology methods (reviewed in Gerhold et al., 2015). As proposed to date, these methods are oversimplified and are biased due to inaccurate taxonomic annotations of intrinsically low-resolution data (Gerhold et al., 2015). In sub-community co-cultures, increased accuracy of functional annotation would enable testing of trait distribution to infer the role of competition or facilitative interactions for example, increasing our understanding of the ecological processes in the microbiome.

## REFERENCES

- Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103
- Amato, S. M., Fazen, C. H., Henry, T. C., Mok, W. W. K., Orman, M. A., Sandvik, E. L., et al. (2014). The role of metabolism in bacterial persistence. *Front. Microbiol.* 5:70. doi: 10.3389/fmicb.2014.00070
- Barona-Gómez, F., Cruz-Morales, P., and Noda-García, L. (2012). What can genome-scale metabolic network reconstructions do for prokaryotic systematics? *Antonie van Leeuwenhoek* 101, 35–43. doi: 10.1007/s10482-011-9655-1
- Blodgett, J. A., Zhang, J. K., Yu, X., and Metcalf, W. W. (2015). Conserved biosynthetic pathways for phosalone, bialaphos and newly discovered

Finally, metagenomic sequencing of co-cultures should be coupled with sampling strategies that enable explicit hypothesis of local adaptation and increase the likelihood of identifying genes and their metabolites with specialized functions in the community. For instance, sampling with biological replicates along environmental gradients, contrasting habitats, or within specialized organs or structures (gut, nodules, specialized roots, etc) would be examples of targets that could increase ecological resolution in interpretation of the microbiome interactions.

## CONCLUSION

To increase our understanding of the microbiome's evolutionary processes, and in particular of the interactions of the microbial community and its environment and hosts, phylogenomics can be used to identify functionally important genes, and co-cultures can be used to approximate sub-community interactions. In combination, our understanding of the microbiome community taxa composition and their interactions will be greatly enhanced.

## AUTHOR CONTRIBUTIONS

AC-J and FB-G both equally conceived the ideas for this manuscript, wrote the text, and revised it critically for intellectual content.

## FUNDING

FB-G is funded by Conacyt (grants #179290 and #177568). AC-J is funded by A01, A.C. #0623, and Conacyt #169701.

## ACKNOWLEDGMENTS

We especially thank Pablo Cruz-Morales for providing unpublished data, help on editing figures, and being a sustained source of thought-provoking discussions. We also thank Jose A. Corona for helpful discussions.

- phosphonic acid natural products. *J. Antibiot. (Tokyo)* doi: 10.1038/ja.2015.77 [Epub ahead of print].
- Boon, E., Meehan, C. J., Whidden, C., Wong, D. H. J., Langille, M. G. I., and Beiko, R. G. (2014). Interactions in the microbiome: communities of organisms and communities of genes. *FEMS Microbiol. Rev.* 38, 90–118. doi: 10.1111/1574-6976.12035
- Bordenstein, S. R., and Theis, R. K. (2015). Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol.* 13:e1002226. doi: 10.1371/journal.pbio.1002226
- Carr, R., Shen-Orr, S. S., and Borenstein, E. (2013). Reconstructing the genomic content of microbiome taxa through shotgun metagenomic deconvolution. *PLoS Comput. Biol.* 9:e1003292. doi: 10.1371/journal.pcbi.1003292
- Chai, J., Kora, G., Ahn, H.-T., Hyatt, D., and Pan, C. (2014). Functional phylogenomics analysis of bacteria and archaea using consistent genome

- annotation with UniFam. *BMC Evol. Biol.* 14:207. doi: 10.1186/s12862-014-0207-y
- Charlop-Powers, Z., Owen, J. G., Reddy, B. V. B., Ternei, M. A., Guimarães, D. O., and de Frias, U. A. (2015). Global biogeographic sampling of bacterial secondary metabolism. *Elife* 4, e05048. doi: 10.7554/eLife.05048
- Chiu, J. C., Lee, E. K., Egan, M. G., Sarkar, I. N., Coruzzi, G. M., and DeSalle, R. (2006). OrthologID: automation of genome-scale ortholog identification within a parsimony framework. *Bioinformatics* 22, 699–707. doi: 10.1093/bioinformatics/btk040
- Cibrián-Jaramillo, A., De la Torre-Bárcena, J. E., Lee, E. K., Katari, M. S., Little, D. P., Stevenson, D. W., et al. (2010). Using phylogenomic patterns and gene ontology to identify proteins of importance in plant evolution. *Genome Biol. Evol.* 2, 225–239. doi: 10.1093/gbe/evq012
- Clements, F. E. (1916). *Plant Succession; An Analysis of the Development of Vegetation*. Washington, DC: Carnegie Institution of Washington.
- Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A. IV, Bik, H. M., and Eisen, J. A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2, e243. doi: 10.7717/peerj.243
- D'Onofrio, A., Crawford, J. M., Stewart, E. J., Witt, K., Gavrish, E., Epstein, S., et al. (2010). Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* 17, 254–264. doi: 10.1016/j.chembiol.2010.02.010
- Fayle, T. M., Turner, E. C., Basset, Y., Ewers, R. M., Reynolds, G., and Novotny, V. (2015). Whole-ecosystem experimental manipulations of tropical forests. *Trends Ecol. Evol.* 30, 334–346. doi: 10.1016/j.tree.2015.03.010
- Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, M. A., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2329–E2338. doi: 10.1073/pnas.1319284111
- Gatesy, J. B. R. (2005). Hidden likelihood support in genomic data: can forty-five wrongs make a right? *Syst. Biol.* 54, 483–492. doi: 10.1080/10635150590945368
- Gerhold, P., Cahill, J. F., Winter, M., Bartish, I. V., and Prinzing, A. (2015). Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Funct. Ecol.* 29, 600–614. doi: 10.1111/1365-2435.12425
- Hittinger, C. T., Rokas, A., and Carroll, S. B. (2004). Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14144–14149. doi: 10.1073/pnas.0404319101
- Hoefer, B. C., Gorzelnik, K. V., Yang, J. Y., Hendricks, N., Dorrestein, P. C., and Straight, P. D. (2012). Enzymatic resistance to the lipopeptide surfactin as identified through imaging mass spectrometry of bacterial competition. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13082–13087. doi: 10.1073/pnas.1205586109
- Hug, L., Beiko, R., Rowe, A., Richardson, R., and Edwards, E. (2012). Comparative metagenomics of three *Dehalococcoides*-containing enrichment cultures: the role of the non-dechlorinating community. *BMC Genomics* 13:327. doi: 10.1186/1471-2164-13-327
- Kearney, M. (2006). Habitat, environment and niche: what are we modelling? *Oikos* 115, 186–191. doi: 10.1111/j.2006.0030-1299.14908.x
- Klitgord, N., and Segrè, D. (2011). Ecosystems biology of microbial metabolism. *Curr. Opin. Biotechnol.* 22, 541–546. doi: 10.1016/j.copbio.2011.04.018
- Konopka, A. (2009). What is microbial community ecology[quest]. *ISME J.* 3, 1223–1230. doi: 10.1038/ismej.2009.88
- Langenheder, S., Lindström, E. S., and Tranvik, L. J. (2006). Structure and function of bacterial communities emerging from different sources under identical conditions. *Appl. Environ. Microbiol.* 72, 212–220. doi: 10.1128/AEM.72.1.212-220.2006
- Lee, E. K., Cibrián-Jaramillo, A., Kolokotronis, S.-O., Katari, M. S., Stamatakis, A., and Ott, M. (2011). A functional phylogenomic view of the seed plants. *PLoS Genet.* 7:e1002411. doi: 10.1371/journal.pgen.1002411
- Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., et al. (2015). A new antibiotic kills pathogens without detectable resistance. *Nature* 517, 455–459. doi: 10.1038/nature14098
- Lok, C. (2015). Mining the microbial dark matter. *Nature* 522, 270–273. doi: 10.1038/522270a
- Luo, Y., Li, B.-Z., Liu, D., Zhang, L., Chen, Y., and Jia, B. (2015). Engineered biosynthesis of natural products in heterologous hosts. *Chem. Soc. Rev.* 44, 5265–5290. doi: 10.1039/c5cs00025d
- Mackelprang, R., Waldrop, M. P., DeAngelis, K. M., David, M. M., Chavarria, K. L., Blazewicz, S. J., et al. (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 480, 368–371. doi: 10.1038/nature10576
- Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39(Suppl. 2), W339–W346. doi: 10.1093/nar/gkr466
- Mills, L. S., Soulé, M. E., and Doak, D. F. (1993). The keystone-species concept in ecology and conservation. *Bioscience* 43, 219–224. doi: 10.1111/e1.12014
- Narechania, A., Baker, R. H., Sit, R., Kolokotronis, S.-O., DeSalle, R., and Planet, P. J. (2012). Random Addition Concatenation Analysis: a novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria. *Genome Biol. Evol.* 4, 30–43. doi: 10.1093/gbe/evr121
- Nguyen, N.-P., Mirarab, S., Liu, B., Pop, M., and Warnow, T. (2014). TIPP: taxonomic identification and phylogenetic profiling. *Bioinformatics* 30, 3548–3555. doi: 10.1093/bioinformatics/btu721
- Noda-García, L., Juárez-Vázquez, A. L., Ávila-Arcos, M. C., Verdusco-Castro, E. A., Montero-Morán, G., and Gaytán, P. (2015). Insights into the evolution of enzyme substrate promiscuity after the discovery of ( $\beta\alpha$ )(8) isomerase evolutionary intermediates from a diverse metagenome. *BMC Evol. Biol.* 15:107. doi: 10.1186/s12862-015-0378-1
- Pagaling, E., Strathdee, F., Spears, S. B., Cates, M. E., Allen, R. J., and Free, A. (2014). Community history affects the predictability of microbial ecosystem development. *ISME J.* 8, 19–30. doi: 10.1038/ismej.2013.150
- Rokas, A., Williams, B. L., King, N., and Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804. doi: 10.1038/nature02053
- Rosenberg, E., Sharon, G., Atad, I., and Zilber-Rosenberg, I. (2010). The evolution of animals and plants via symbiosis with microorganisms. *Environ. Microbiol. Rep.* 2, 500–506. doi: 10.1111/j.1758-2229.2010.00177.x
- Salichos, L., Stamatakis, A., and Rokas, A. (2014). Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Mol. Biol. Evol.* 31, 1261–1271. doi: 10.1093/molbev/msu061
- Sharon, I., Morowitz, M. J., Thomas, B. C., Costello, E. K., Relman, D. A., and Banfield, J. F. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res.* 23, 111–120. doi: 10.1101/gr.142315.112
- Shi, T., and Falkowski, P. G. (2008). Genome evolution in cyanobacteria: the stable core and the variable shell. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2510–2515. doi: 10.1073/pnas.0711165105
- Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Kultima, J. R., et al. (2013). Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* 10, 1196–1199. doi: 10.1038/nmeth.2693
- Traxler, M. F., and Kolter, R. (2015). Natural products in soil microbe interactions and evolution. *Nat. Prod. Rep.* 32, 956–970. doi: 10.1039/c5np00013k
- van Tonder, A. J., Mistry, S., Bray, J. E., Hill, D. M. C., Cody, A. J., Farmer, C. L., et al. (2014). Defining the estimated core genome of bacterial populations using a bayesian decision model. *PLoS Comput. Biol.* 10:e1003788. doi: 10.1371/journal.pcbi.1003788
- Waldor, M. K., Tyson, G., Borenstein, E., Ochman, H., Moeller, A., Finlay, B. B., et al. (2015). Where next for microbiome research? *PLoS Biol.* 13:e1002050. doi: 10.1371/journal.pbio.1002050
- Wilson, M. C., Mori, T., Ruckert, C., Uria, A. R., Helf, M. J., Takada, K., et al. (2014). An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 506, 58–62. doi: 10.1038/nature12959
- Wu, C., Zacchetti, B., Ram, A. F. J., van Wezel, G. P., Claessen, D., and Choi, H. Y. (2015). Expanding the chemical space for natural products by *Aspergillus-Streptomyces* co-cultivation and biotransformation. *Sci. Rep.* 5, 10868. doi: 10.1038/srep10868
- Wu, D., Wu, M., Halpern, A., Rusch, D. B., Yooshep, S., Frazier, M., et al. (2011). Stalking the Fourth Domain in Metagenomic Data: searching for, discovering,

and interpreting novel, deep branches in marker gene phylogenetic trees. *PLoS ONE* 6:e18011. doi: 10.1371/journal.pone.0018011

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Cibrián-Jaramillo and Barona-Gómez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.